

On the regularity of iterated hairpin completion of a single word

Lila Kari, Steffen Kopecki, and Shinnosuke Seki

January 19, 2013

Abstract

Hairpin completion is an abstract operation modeling a DNA bio-operation which receives as input a DNA strand $w = x\alpha y\bar{\alpha}$, and outputs $w' = x\alpha y\bar{\alpha}\bar{x}$, where \bar{x} denotes the Watson-Crick complement of x . In this paper, we focus on the problem of finding conditions under which the iterated hairpin completion of a given word is regular. According to the numbers of words α and $\bar{\alpha}$ that initiate hairpin completion and how they are scattered, we classify the set of all words w . For some basic classes of words w containing small numbers of occurrences of α and $\bar{\alpha}$, we prove that the iterated hairpin completion of w is regular. For other classes with higher numbers of occurrences of α and $\bar{\alpha}$, we prove a necessary and sufficient condition for the iterated hairpin completion of a word in these classes to be regular.

1 Introduction

A DNA strand can be abstractly viewed as a word over the alphabet $\{A, C, G, T\}$, where in A is Watson-Crick complementary to T and C to G , and two complementary DNA single strands of opposite orientation bind together to form a double DNA strand (intermolecular structure). Also, if subwords of a DNA strand are complementary, the strand may bind to itself forming intramolecular structures such as stem-loops, also known more commonly as *hairpins* (Figure 1 (2)). Hairpins can be a building block of a larger-scale structure of RNA strands, and play a role in determining various chemical and thermodynamical properties (stability, structures, functions) of the structure, and make significant contributions to the genetic information processing as illustrated in their function as a stopper for messenger RNA (mRNA) transcription. A CG-rich sequence of an mRNA folds into its Watson-Crick complement on the RNA and forms a stable hairpin. Transcription of the mRNA is terminated when RNA polymerase reaches the hairpin. At that time, *nusA* protein bound to the polymerase interacts with the hairpin and takes the polymerase off the mRNA. This hairpin-driven mechanism is called *intrinsic termination* [23]. As such, hairpins tend to interfere with reactions, and therefore were given the cold shoulder by DNA computing

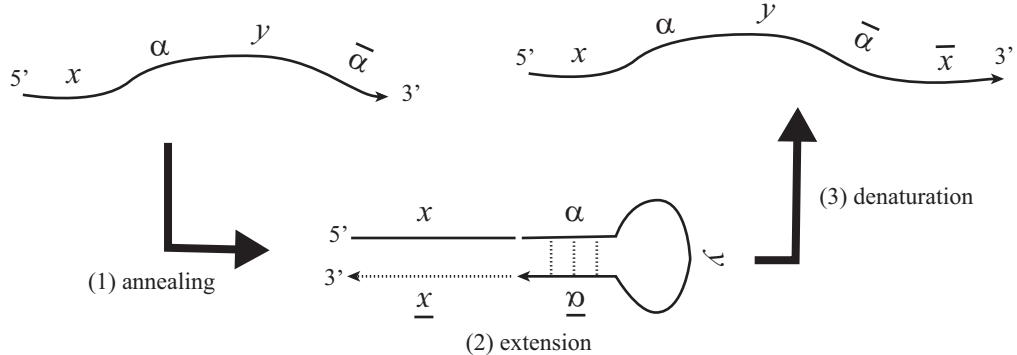


Figure 1: Hairpin completion by polymerase chain reaction [7, 20]. The operation input is $x\alpha y\bar{\alpha}$, the output is $x\alpha y\bar{\alpha}\bar{x}$, and the primer is α .

experimentalists. See [1, 2, 9, 10, 12, 19] about this problem and about some of the “good” designs of DNA strands that are free of hairpins.

Hairpin is not a foe to all DNA computing experiments; many molecular computing machineries have been proposed which make good use of hairpins. Such hairpin-driven systems include DNA RAM [11, 21, 22] and Whiplash PCR [7, 20]. In particular, Whiplash PCR features a self-directed polymerase chain reaction (PCR) of DNA strand, which practically motivates the investigation of a formal language operation called *hairpin completion*. Hairpin completion proceeds as follows (Figure 1): Starting from a DNA strand $w = x\alpha y\bar{\alpha}$, a segment $\bar{\alpha}$ at the 3'-end of w binds to its Watson-Crick complementary strand α on the strand (annealing). A polymerase chain reaction then extends w at its 3'-end in the $5' \rightarrow 3'$ direction so as to generate the strand $x\alpha y\bar{\alpha}\bar{x}$ (let us call α and $\bar{\alpha}$ that bind with each other to initiate this PCR reaction *primers*). Despite the intrinsic $5' \rightarrow 3'$ polarity of polymerases, a mechanism exists to make polymerase reaction work in the $3' \rightarrow 5'$ direction (Okazaki fragment [18]).

As an abstract model of the above-mentioned self-directed PCR, Cheptea, Martín-Vide, and Mitrana proposed the hairpin completion in [3], and since then this abstract operation has been studied on its algorithmic and formal linguistic aspects [5, 15, 16, 17] together with its variant called *bounded hairpin completion* [8, 14], where the length of extension in one operation is bounded by a constant. Ito et al. [8] and Kopecki [14] proved that all classes in the Chomsky Hierarchy are closed under iterated bounded hairpin completion. In contrast, the class of regular languages was proved not to be closed under iterated (unbounded) hairpin completion [3], and a surprising fact is that iterated hairpin completion of a word can be non-regular [14]. In this paper, we focus on a problem proposed by Kopecki in [14]; is it decidable whether the iterated hairpin completion of a given word is regular? The iterated hairpin completion of a singleton language

(a word) is known to be in NL [3], but can be non-regular as shown in the following example.

Example 1. Let $\alpha = a^k$ and $w = abaca\bar{a}\bar{d}\bar{a}$, where $a, \bar{a}, b, \bar{b}, c, \bar{c}, d, \bar{d}$ are all distinct letters. Then the intersection of the iterated hairpin completion of w with $(abac(ab)^+ad)^2abaca\bar{a}\bar{d}\bar{a}(\bar{b}\bar{a})^+\bar{c}\bar{a}\bar{b}\bar{a}$ is $\{(\alpha bac(\alpha b)^i\alpha d)^2\alpha bac a\bar{a}\bar{d}\bar{a}(\bar{b}\bar{a})^i\bar{c}\bar{a}\bar{b}\bar{a} \mid i \geq 1\}$. This intersection is not context-free, and neither is the iterated hairpin completion.

In this paper, we give a partial answer to the regularity-test decidability problem. We focus our attention on the number of primers a given word contains as its factors and on how these primers are scattered over the given word. All the words are classified in accordance with these two criteria, and for some basic classes, we give a necessary and sufficient condition for the iterated hairpin completion of a word in the class to be regular.

2 Preliminaries

Let Σ be an alphabet, Σ^* be the set of all words over Σ , and for an integer $k \geq 0$, Σ^k be the set of all words of length k over Σ . The word of length 0 is called the empty word, denoted by λ , and let $\Sigma^+ = \Sigma^* \setminus \{\lambda\}$. A subset of Σ^* is called a language over Σ . For a word $w \in \Sigma^*$, we employ the notation w when we mean the word as well as the singleton language $\{w\}$ unless confusion arises. For a language $L \subseteq \Sigma^*$, we denote by L^* the set $\{w_1 \cdots w_n \mid n \geq 0, w_1, \dots, w_n \in L\}$.

We equip Σ with a function $\bar{} : \Sigma \rightarrow \Sigma$ satisfying $\forall a \in \Sigma, \bar{\bar{a}} = a$; such a function is called an *involution*. This involution $\bar{}$ is naturally extended to words as: for $a_1, a_2, \dots, a_n \in \Sigma$, $\overline{a_1 a_2 \cdots a_n} = \overline{a_n} \cdots \overline{a_2} \overline{a_1}$. For example, over the 4-letter alphabet $\Delta = \{A, C, G, T\}$, if we define an involution $d : \Delta \rightarrow \Delta$ as $d(A) = T$ and $d(C) = G$, then d , being thus extended, maps the Watson strand of a complete DNA double strand into its Crick strand. The involution d is called the Watson-Crick involution [13]. For a word $w \in \Sigma^*$, we call \overline{w} the *complement* of w , being inspired by this application. A word $w \in \Sigma^*$ is called a *pseudo-palindrome* if $w = \overline{w}$. For a language $L \subseteq \Sigma^*$, $\overline{L} = \{\overline{w} \mid w \in L\}$.

For words $u, w \in \Sigma^*$, if $w = xuy$ holds for some words $x, y \in \Sigma^*$, then u is called a *factor* of w ; a factor that is distinct from w is said to be *proper*. If the equation holds with $x = \lambda$ ($y = \lambda$), then the factor u is especially called a *prefix* (resp. a *suffix*) of w . The prefix relation can be regarded as a partial order \leq_p over Σ^* ; $u \leq_p w$ means that u is a prefix of w . Analogously, by $w \geq_s v$ we mean that v is a suffix of w . For a word $w \in \Sigma^*$ and a language $L \subseteq \Sigma^*$, a factor v of w is *minimal with respect to L* if $v \in L$ and none of the proper factors of v is in L .

A nonempty word $w \in \Sigma^+$ is *primitive* if $w = x^i$ implies $i = 1$ for any nonempty word $x \in \Sigma^+$. It is well-known that for any nonempty word w , there exists a unique primitive word u with $w \in u^+$. Such u is called the *primitive root* of w and denoted by $\rho(w)$. Two words $x, y \in \Sigma^*$ *commute* if $xy = yx$, and this is known to be equivalent to $\rho(x) = \rho(y)$. See [4] for details of primitivity and commutativity of words and related results.

Now we introduce the operation investigated in this paper, that is, hairpin completion, and define it formally. Imagine that we have a DNA sequence $5' - \text{CAATCGTATGAT} - 3'$. The suffix **GAT** can find its d -image as a factor **ATC** on this sequence. Hence, this DNA sequence may bend over into a hairpin form by **GAT** binding with **ATC**. This formation of hairpin structure leaves **CA** as a free sticky-end, and DNA polymerase converts it into the complete double strand by extending its 3'-end by **TG** = $d(\text{CA})$. This exemplifies the mechanism of hairpin completion. We call two words whose thus binding initiate hairpin completion *primers*. In the above example, **GAT** and **ATC** are primers.

Let k be a constant that is assumed to be the length of a primer. Throughout this paper, we will not use the notation ‘ k ’ for any other purpose. Let $\alpha \in \Sigma^k$ be a primer. If a given word $w \in \Sigma^*$ has a factorization $u\alpha v\bar{\alpha}$ for some $u, v \in \Sigma^*$ and $\alpha \in \Sigma^k$, then its *right hairpin completion with respect to α* results in the word $u\alpha v\bar{\alpha}u$. As long as α is clear from context, this operation is simply called (*single-primer*) *right hairpin completion*. By $w \rightarrow_{\mathcal{RH}_\alpha} w'$, or by $w \rightarrow_{\mathcal{RH}} w'$, we mean that w' can be obtained from w by right hairpin completion (with respect to α). The *left hairpin completion* is defined analogously as an operation to derive $u'\alpha v'\bar{\alpha}u'$ from $\alpha v'\bar{\alpha}u'$, and the relation $\rightarrow_{\mathcal{LH}_\alpha}$ is naturally introduced. By $\rightarrow_{\mathcal{LH}}^*$ and $\rightarrow_{\mathcal{RH}}^*$, we denote the reflexive transitive closure of $\rightarrow_{\mathcal{LH}}$ and that of $\rightarrow_{\mathcal{RH}}$, respectively. The relation $\rightarrow_{\mathcal{H}}$ is defined as the union of $\rightarrow_{\mathcal{LH}}$ and $\rightarrow_{\mathcal{RH}}$.

For a given language $L \subseteq \Sigma^*$, we define the set of words obtained by left hairpin completion from L , and the set of words obtained by iterated left hairpin completion from L , respectively, as follows:

$$\mathcal{LH}_\alpha(L) = \{w' \mid \exists w \in L, w \rightarrow_{\mathcal{LH}_\alpha} w'\}, \quad \mathcal{LH}_\alpha^*(L) = \{w' \mid \exists w \in L, w \rightarrow_{\mathcal{LH}_\alpha}^* w'\}.$$

Analogously, $\mathcal{RH}_\alpha(L)$ and $\mathcal{RH}_\alpha^*(L)$ are defined based on $\rightarrow_{\mathcal{RH}}$ and $\rightarrow_{\mathcal{RH}}^*$, and $\mathcal{H}_\alpha(L)$ and $\mathcal{H}_\alpha^*(L)$ are defined based on $\rightarrow_{\mathcal{H}}$ and $\rightarrow_{\mathcal{H}}^*$.

Proposition 1. *For a word $w \in \Sigma^*$, $\mathcal{RH}_k^*(w) = \overline{\mathcal{LH}_k^*(\overline{w})}$.*

3 Word structures relevant to the power of iterated hairpin completion

In this section, we describe several structural properties of a word w that will be relevant for the characterization of its iterated hairpin completion $\mathcal{H}_\alpha^*(w)$, where $\alpha \in \Sigma^k$ is a fixed parameter.

A word $u \in \Sigma^*$ is called an α -*prefix* of a word $w \in \Sigma^*$ if $w = u\alpha x$ for some word $x \in \Sigma^*$. In a similar manner, a word $v \in \Sigma^*$ is an $\bar{\alpha}$ -*suffix* of w if $w = y\bar{\alpha}v$ for some $y \in \Sigma^*$. If $w = y\bar{\alpha}v$ begins with α , then this prefix can bind with the occurrence of $\bar{\alpha}$ (unless they overlap with each other), and left hairpin completion results in $\bar{v}w$. By $\text{Pref}_\alpha(w)$ and $\text{Suff}_{\bar{\alpha}}(w)$, we denote the set of all α -prefixes and that of all $\bar{\alpha}$ -suffixes of w , respectively. One can easily observe that $\text{Suff}_{\bar{\alpha}}(w) = \overline{\text{Pref}_\alpha(\overline{w})}$. Throughout this paper, we let $\text{Pref}_\alpha(w) = \{u_1, \dots, u_m\}$

and $\text{Suff}_{\overline{\alpha}}(w) = \{\overline{v_1}, \dots, \overline{v_n}\}$ for some $m, n \geq 0$. It will be convenient to assume that these α -prefixes are sorted in the ascending order of their length. Likewise, we assume that $|\overline{v_1}| < |\overline{v_2}| < \dots < |\overline{v_n}|$.

Our investigation on the properties of α -prefix and $\overline{\alpha}$ -suffix of word begins with a basic observation.

Proposition 2. *For a word $w \in \alpha\Sigma^*$, the following statements hold:*

1. *for any $u \in \text{Pref}_\alpha(w)$, $\alpha \leq_p u\alpha$;*
2. *for any $x_1, \dots, x_n \in \text{Pref}_\alpha(w)$, $\alpha \leq_p x_1 \dots x_n \alpha$;*

Proof. The first statement derives directly from the definition of α -prefix. For the second one, induction on n works. Due to the first statement, $\alpha \leq_p x_n \alpha$ so that proving $\alpha \leq_p x_1 \dots x_{n-1} x_n \alpha$ is reduced to proving $\alpha \leq_p x_1 \dots x_{n-1} \alpha$. \square

From this proposition, we can easily deduce that for a word $w \in \Sigma^* \overline{\alpha}$ and $\overline{y_1}, \dots, \overline{y_t} \in \text{Suff}_{\overline{\alpha}}(w)$, $\overline{\alpha} \overline{y_1} \dots \overline{y_t} \geq_s \overline{\alpha}$, which means $\alpha \leq_p y_t \dots y_1 \alpha$. This deepens the above observation further as follows.

Corollary 1. *For a word $w \in \alpha\Sigma^* \cap \Sigma^* \overline{\alpha}$, any word in $(\text{Pref}_\alpha(w) \cup \overline{\text{Suff}_{\overline{\alpha}}(w)})^* \alpha$ has α as its prefix.*

Due to the second statement of Proposition 2, $\alpha \leq_p x_1 \alpha \leq_p x_1 x_2 \alpha \leq_p \dots \leq_p x_1 x_2 \dots x_s \alpha$ holds for α -prefixes $x_1, \dots, x_s \in \text{Pref}_\alpha(w)$. Hence, from a word $x_1 x_2 \dots x_s \alpha w' \overline{\alpha}$, one-step right hairpin completion can produce at least the words $x_1 x_2 \dots x_s \alpha w' \overline{\alpha} \{\lambda, \overline{x_1}, \overline{x_1 x_2}, \dots, \overline{x_1 x_2 \dots x_s}\}$.¹ Now, if we know that one-step hairpin completion extends the word to the right by \overline{u} , what can we say about the word u ? Firstly, as long as $|u| \leq |x_1 \dots x_s|$, we can say that $u\alpha \leq_p x_1 \dots x_s \alpha$ by definition of hairpin completion. Moreover, Corollary 1 enables us to find $0 \leq i < s$ such that $|x_1 \dots x_i| < |u| \leq |x_1 \dots x_{i+1}|$. Then, one can let $u = x_1 \dots x_i z$ for some prefix z of x_{i+1} . Since $z\alpha \leq_p x_{i+1} \alpha \leq_p w$, z is an α -prefix of w that is properly shorter than x_{i+1} . By defining $\text{ind}(x_{i+1})$ to be the index satisfying $u_{\text{ind}(x_{i+1})} = x_{i+1}$, we have $z \in \{u_1, \dots, u_{\text{ind}(x_{i+1})-1}\}$; recall that elements of $\text{Pref}_\alpha(w)$ is sorted with respect to their length. The above argument is summarized by the next lemma.

Lemma 1. *Let $x_1, \dots, x_s \in \text{Pref}_\alpha(w)$. If a word u satisfies $u\alpha \leq_p x_1 \dots x_s \alpha$, then there exists an integer $0 \leq i < s$ such that $u = x_1 \dots x_i z$ for some $z \in \{u_1, \dots, u_{\text{ind}(x_{i+1})-1}\}$.*

A more natural setting is to assume that each of x_1, \dots, x_s is either an element of $\text{Pref}_\alpha(w)$ or an element of $\overline{\text{Suff}_{\overline{\alpha}}(w)}$ because, by left hairpin completion, the complement of a $\overline{\alpha}$ -suffix of w can be produced to the left of w . We need to generalize the function ind by extending its domain as follows: for $x_i \in \overline{\text{Suff}_{\overline{\alpha}}(w)}$, $\text{ind}(x_i) = j$ if $x_i = v_j$. Note that this generalized ind is not a function any more in cases when $\text{Pref}_\alpha(w) \cap \overline{\text{Suff}_{\overline{\alpha}}(w)} \neq \emptyset$, but this will not cause any problem in this paper.

¹ $\overline{x_1 x_2 \dots x_s} = \overline{x_s} \dots \overline{x_2} \overline{x_1}$.

Lemma 2. Let $x_1, \dots, x_s \in \text{Pref}_\alpha(w) \cup \overline{\text{Suff}_\alpha(w)}$. If a word u satisfies $u\alpha \leq_p x_1 \cdots x_t \alpha$, then there exists an integer $0 \leq i < s$ such that $u = x_1 \cdots x_i z$, where

$$\begin{cases} z \in \{u_1, \dots, u_{\text{ind}(x_{i+1})-1}\} & \text{if } x_{i+1} \in \text{Pref}_\alpha(w); \\ z \in \{v_1, \dots, v_{\text{ind}(x_{i+1})-1}\} & \text{if } x_{i+1} \in \overline{\text{Suff}_\alpha(w)}. \end{cases}$$

Proof. As done previously, we can find $0 \leq i < s$ and a nonempty word $z \in \Sigma^+$ satisfying $u = x_1 \cdots x_i z$ and $z\alpha \leq_p x_{i+1}\alpha$. Since this prefix relation can be rewritten as $\bar{\alpha}x_{i+1} \geq_s \bar{\alpha}\bar{z}$, if \bar{x}_{i+1} is an $\bar{\alpha}$ -suffix of w , so is \bar{z} . The case when $x_{i+1} \in \text{Pref}_\alpha(w)$ is clear from the previous argument. \square

Having considered prefix relations among α -prefixes and $\bar{\alpha}$ -suffixes of a word, now we proceed our study to more general factor relationships among them.

Lemma 3. If $u_j\alpha \geq_s u_i\alpha$ for some integers $2 \leq i \leq j \leq m$, then $u_j \in \{u_1, u_2, \dots, u_{j-1}\}u_i$.

Proof. We can let $xu_i\alpha = u_j\alpha$ for some $x \in \Sigma^*$. Combining this with Proposition 2, we have $x\alpha \leq_p u_j\alpha$ so that $x \in \text{Pref}_\alpha(w)$. Since $|x| < |u_j|$, x is in $\{u_1, u_2, \dots, u_{j-1}\}$. \square

Lemma 4. If $v_2\alpha$ is a factor of $u_2\alpha$, then $u_2 = v_2$.

Proof. Let $u_2\alpha = xv_2\alpha y$ for some $x, y \in \Sigma^*$. Unless $y = \lambda$, $xv_2\alpha \leq_p u_2\alpha$ would be a nonempty α -prefix of w that is properly shorter than u_2 , and causes a contradiction. Thus, y must be empty so that $u_2\alpha = xv_2\alpha$. Now, Lemma 3 leads us to $x = \lambda$. \square

Finally, let us introduce interesting results that illustrate the close relationship between α -prefixes, commutativity, and primitivity, essential notions in combinatorics on words.

Lemma 5. Let $w \in \alpha\Sigma^*$ and $u \in \text{Pref}_\alpha(w)$. Then $\rho(u), \rho(u)^2, \dots, \rho(u)^{|u|/|\rho(u)|} \in \text{Pref}_\alpha(w)$.

Proof. Due to the first statement of Proposition 2, $u \in \text{Pref}_\alpha(w)$ enables us to let $\alpha y = u\alpha$ for some $y \in \Sigma^+$. Its solution is well-known to be $u = (st)^n$ and $\alpha = (st)^i s$ for some $i \geq 0$ and $s, t \in \Sigma^*$ such that $\rho(u) = st$. Hence, $u\alpha = (st)^{i+n}s = \rho(u)\alpha(ts)^{n-1} = \rho(u)^2\alpha(ts)^{n-2} = \dots = \rho(u)^n\alpha$. \square

An immediate implication of this lemma is that the shortest nonempty α -prefix of a word that begins with α must be primitive. We should make one more step forward. Imagine that a word w has an α -prefix u . If $w \rightarrow_{\mathcal{RH}} w\bar{u}$ is possible, then $w \rightarrow_{\mathcal{RH}} w\rho(u)$ is also possible. Thus, repeating the extension of w to the right by $\rho(u)$ $|u|/|\rho(u)|$ times amounts to extending w by \bar{u} once. In other words, the process to extend a word by \bar{u} is not essential unless u is primitive because it can be always simulated by multiple processes to extend a word by $\rho(u)$.

The next lemma proves that all nonempty α -prefixes of length at most $|\alpha|$ commute with each other, and hence, only the shortest one is essential in the above sense.

Lemma 6. *For nonempty words $x_1, x_2 \in \Sigma^+$, if $\alpha \leq_p x_1\alpha \leq_p x_2\alpha$ and $|x_2| \leq |\alpha|$ hold, then $\rho(x_1) = \rho(x_2)$.*

Proof. If $|x_1| = |x_2|$, then the prefix relation immediately gives $x_1 = x_2$, and the conclusion of this lemma is trivial. Hence, we assume $|x_1| < |x_2|$. Combining $|x_1| \leq |\alpha|$ with $\alpha \leq_p x_1\alpha$, we can deduce that the word $x_1\alpha$ has a period $|x_1|$. Likewise, $x_2\alpha$ has a period $|x_2|$, and hence, $x_1\alpha$ also has this period. As a result, $x_1\alpha$ has two periods $|x_1|, |x_2|$, and moreover it is of length at least the sum of these periods. Thus, Fine and Wilf's theorem [4, 6] leads us to the conclusion of this lemma. \square

3.1 Non-crossing words and their properties

A word $w_0 \in \Sigma^*$ is an (m, n) - α -word, or simply an (m, n) -word when α is clear from the context, if $|\text{Pref}_\alpha(w_0)| = m$ and $|\text{Suff}_{\bar{\alpha}}(w_0)| = n$. Informally speaking, an (m, n) -word is a word on which α occurs m times and $\bar{\alpha}$ does n times. For a pseudo-palindromic α ($\alpha = \bar{\alpha}$), we regard an occurrence of α also as that of $\bar{\alpha}$, and as such, any word is an (m, m) -word for some $m \geq 0$.

We say that w_0 is *non- α -crossing* if the rightmost occurrence of α precedes the leftmost one of $\bar{\alpha}$ on w_0 . When α is understood from the context, we simply say that w_0 is non-crossing. Otherwise, the word is α -crossing or crossing. Note that if $\alpha = \bar{\alpha}$, then for a word w which is either a $(0, 0)$ -word or $(1, 1)$ -word, $\mathcal{H}_\alpha^*(w) = \{w\}$, and otherwise (w is an (m, m) -word for some $m \geq 2$), w can be considered crossing. Thus, whenever the non- α -crossing word is concerned, we assume that $\alpha \neq \bar{\alpha}$. The definition of a word being non- α -crossing does not force the word to begin with α or end with $\bar{\alpha}$. However, it is not until α is a primer that this notion becomes useful in our work. Thus, the word should be in either $\alpha\Sigma^*$ or $\Sigma^*\bar{\alpha}$. Actually, in the rest of this paper, we assume both of these conditions and consider only *single-primer iterated hairpin completion*; thus, we can assume that $w_0 \in \alpha\Sigma^* \cap \Sigma^*\bar{\alpha}$. As let previously, elements of $\text{Pref}_\alpha(w_0)$ are denoted by u_1, \dots, u_m , those of $\text{Suff}_{\bar{\alpha}}(w_0)$ by $\bar{v}_1, \dots, \bar{v}_n$, and they are sorted so that this assumption imposes $u_1 = v_1 = \lambda$.

Our main focus lies on the characterization of non-crossing words whose iterated hairpin completion is regular in terms of combinatorics on words. Thus, in this subsection, we prove some combinatorial properties of non-crossing words. Let us begin with an easy observation about the longest α -prefix and $\bar{\alpha}$ -suffix of w_0 .

Proposition 3. *$u_m = v_n$ if and only if $m = n$ and for all $1 \leq i \leq m$, $u_i = v_i$.*

Next, we will see that one-step hairpin completion can extend w_0 to the left by any of v_1, \dots, v_{n-1} or to the right by any of $\bar{u}_1, \dots, \bar{u}_{m-1}$ due to the following lemma.

Lemma 7. *Let $w_0 \in \alpha\Sigma^* \cap \Sigma^*\bar{\alpha}$ be a non-crossing word with $\text{Pref}_\alpha(w_0) = \{u_1, \dots, u_m\}$ and $\text{Suff}_{\bar{\alpha}}(w_0) = \{\bar{v}_1, \dots, \bar{v}_n\}$. Then $|u_{m-1}| + |v_n| + 2|\alpha| < |w_0|$.*

Proof. Suppose that this inequality did not hold. Being non-crossing, w_0 can be written as $w_0 = u_{m-1}w\bar{v}_n$ for some $w \in \alpha\Sigma^* \cap \Sigma^*\bar{\alpha}$ with $|w| \leq 2|\alpha|$. Hence, $w = \bar{w}$. Let x be a nonempty word satisfying $u_m = u_{m-1}x$. Since w_0 is non-crossing, $u_m\alpha \leq_p u_{m-1}w$ must hold, from which we have $x\alpha \leq_p w$. Combining this with $w = \bar{w}$ enables us to find an $\bar{\alpha}$ -suffix $\bar{x}\bar{v}_n$ of w_0 , but this would be longer than the longest $\bar{\alpha}$ -suffix of w_0 , a contradiction. \square

This lemma does not rule out the possibility that w_0 cannot be extended to the right by \bar{u}_m by hairpin completion because the rightmost occurrence of α might overlap with the suffix $\bar{\alpha}$. The analogous argument is valid for v_n and left hairpin completion. However, Lemma 7 leads us to one important corollary on non-crossing (m, n) -words for $m, n \geq 2$ that hairpin completion can extend w_0 to the right by the complement of any of its α -prefix and to the left by the complement of any of its $\bar{\alpha}$ -suffix.

Corollary 2. *Let $w_0 \in \alpha\Sigma^* \cap \Sigma^*\bar{\alpha}$ be a non-crossing (m, n) -word with $m, n \geq 2$. Then $\mathcal{H}_\alpha(w_0) = \{w_0\} \cup \{v_2, \dots, v_m\}w_0 \cup w_0\{\bar{u}_2, \dots, \bar{u}_n\}$.*

Any word obtained from a non-crossing word by hairpin completion is non-crossing. Though being easily confirmed, this closure property forms the foundation of our discussions in this paper.

Proposition 4. *Let $\alpha \in \Sigma^k$ with $\alpha \neq \bar{\alpha}$, and $w_0 \in \alpha\Sigma^* \cap \Sigma^*\bar{\alpha}$ be a non-crossing word. Then any word in $\mathcal{H}_\alpha^*(w_0)$ is non-crossing.*

We conclude this section with a characterization of a non- α -crossing word in terms of minimal factors with respect to the language $\alpha\Sigma^* \cap \Sigma^*\bar{\alpha}$. With Proposition 4, this characterization will bring a unique factorization theorem (Theorem 1) of any word w in $\mathcal{H}_\alpha^*(w_0)$ as $w = xw_0y$ for some words x, y .

Lemma 8. *Let $\alpha \in \Sigma^k$ with $\alpha \neq \bar{\alpha}$. A word $w_0 \in \alpha\Sigma^* \cap \Sigma^*\bar{\alpha}$ is non-crossing if and only if it contains exactly one minimal factor v from $\alpha\Sigma^* \cap \Sigma^*\bar{\alpha}$.*

Proof. Let us consider the contrapositive of the converse implication. So, if w_0 is crossing, then we can find an occurrence of $\bar{\alpha}$ (let us denote it by $\bar{\alpha}_0$) which precedes an occurrence of α (α_1). $\bar{\alpha}_0$ is guaranteed to be preceded by another occurrence of α (α_2) because w_0 begins with α . Thus, the factor of w_0 that spans from α_2 to $\bar{\alpha}_0$ is a minimal factor from $\alpha\Sigma^* \cap \Sigma^*\bar{\alpha}$. By the same token, the factor of w_0 that spans from α_2 to its right adjacent occurrence of $\bar{\alpha}$ becomes another minimal factor.

In order to prove the direct implication, suppose that w_0 contains two minimal factors from $\alpha\Sigma^* \cap \Sigma^*\bar{\alpha}$. These two factors must overlap with each other because otherwise the suffix $\bar{\alpha}$ of the first factor precedes the prefix α of the second one and w would be crossing. However, if they overlap, then the overlapped part would be in $\alpha\Sigma^* \cap \Sigma^*\bar{\alpha}$, and this contradicts the minimality of the two factors. \square

Theorem 1. *Let $\alpha \in \Sigma^k$ with $\alpha \neq \bar{\alpha}$, and $w_0 \in \alpha\Sigma^* \cap \Sigma^*\bar{\alpha}$ be a non-crossing word. On any word in $\mathcal{H}_\alpha^*(w_0)$, w_0 occurs exactly once as a factor.*

Proof. From the two facts that any word in $\mathcal{H}_\alpha^*(w_0)$ is non-crossing (Proposition 4) and that these words contain at least one occurrence of w_0 as a factor by definition of hairpin completion, we can reach this conclusion. \square

4 Iterated hairpin completion of non-crossing words

This section contains the main contribution of this paper: characterizations of the regularity of iterated hairpin completion of a non-crossing (m, n) -word $w_0 \in \alpha\Sigma^* \cap \Sigma^*\bar{\alpha}$ (recall that $\alpha \neq \bar{\alpha}$ is assumed). Throughout this section, w_0 is thus assumed with $\text{Pref}_\alpha(w_0) = \{u_1, \dots, u_m\}$ and $\text{Suff}_{\bar{\alpha}}(w_0) = \{\bar{v}_1, \dots, \bar{v}_m\}$.

Let us begin with a proof that one-sided hairpin completion of a non-crossing word is regular (Theorem 2). Then we will show that the iterated hairpin completion of a non-crossing $(m, 1)$ -word for any $m \geq 1$ or $(2, 2)$ -word is always regular (Theorems 3 and 4). Using these results and combinatorial results shown in Section 3, we characterize the set of all non-crossing $(3, 2)$ -words whose iterated hairpin completion is regular, in terms of commutativity (Theorem 5).

Theorem 2. *For a non-crossing word $w_0 \in \alpha\Sigma^* \cap \Sigma^*\bar{\alpha}$, both $\mathcal{LH}_\alpha^*(w_0)$ and $\mathcal{RH}_\alpha^*(w_0)$ are regular.*

Proof. First, we prove the regularity of $\mathcal{RH}_\alpha^*(w_0)$. Let w be an α -prefix of w_0 . A right hairpin completion of w_0 can produce $w_0\bar{w}$. Note that the suffix $\bar{\alpha}\bar{w}$ of this resulting word does not contain α due to the non-crossing assumption on v , and this means that the longest α -prefix of $w_0\bar{w}$ is the same as that of w_0 . Thus, the language $\mathcal{RH}_\alpha^*(w_0)$ can be obtained by iterated bounded hairpin completion from v , and hence, is regular [14].

For the regularity of $\mathcal{LH}_\alpha^*(w_0)$, it suffices to observe that \bar{w}_0 is also non-crossing. Using the result just proved, $\mathcal{RH}_\alpha^*(\bar{w}_0)$ is regular, and according to Proposition 1, $\mathcal{LH}_\alpha^*(w_0) = \mathcal{RH}_\alpha^*(\bar{w}_0)$. Note that the class of regular languages is closed under $^+$. \square

4.1 Iterated hairpin completion of $(m, 1)$ non-crossing words

In this subsection, we consider the case $n = 1$ (w_0 is an $(m, 1)$ -word), and prove that $\mathcal{H}_\alpha^*(w_0)$ is regular. For $m = 1$, it is easy to see that hairpin completion cannot generate any word but w_0 , that is, $\mathcal{H}_\alpha^*(w_0) = \{w_0\}$. Hence, we assume $m \geq 2$.

Lemma 7 means that right hairpin completion can extend w_0 to the right by any of $\bar{u}_1, \bar{u}_2, \dots, \bar{u}_{m-1}$. In contrast, the operation can extend w_0 to the right by \bar{u}_m if and only if $|u_m| + 2|\alpha| + |v_1| \geq |w_0|$, i.e., the α to the right of u_m does not overlap with the suffix $\bar{\alpha}$ of w_0 . As a result, if $m = 2$ but this inequality does not hold, then $\mathcal{H}_\alpha^*(w_0) = \{w_0\}$. Therefore, we can advance our discussion on the assumption that $w_0 \rightarrow_{\mathcal{RH}} w_0\bar{u}_2$ is valid.

Note that $w_0\bar{u}_2$ is a non-crossing $(m, 2)$ -word. Applying Lemma 7 to this word, we can see that $|u_m| + 2|\alpha| < |w_0\bar{u}_2|$. Hence, hairpin completion can

extend $w_0\overline{u_2}$ further to the right by not only by any of $\overline{u_1}, \overline{u_2}, \dots, \overline{u_{m-1}}$ but also by $\overline{u_m}$.

Let us define the following regular language:

$$R_{m1}(w_0) = \{w_0\} \cup \{x_s \cdots x_1 w_0 \overline{y_1} \overline{y_2} \cdots \overline{y_t} \mid \begin{array}{ll} y_1 \in \begin{cases} \{u_1, \dots, u_{m-1}, u_m\} & \text{if } |u_m| + 2|\alpha| \leq |w_0| \\ \{u_1, \dots, u_{m-1}\} & \text{otherwise} \end{cases} \\ s \geq 0, t \geq 1, x_s, \dots, x_1, y_2, \dots, y_t \in \{u_1, \dots, u_m\}, \\ \text{and } \max_{1 \leq i \leq s} \{\text{ind}(x_i)\} \leq \max_{1 \leq j \leq t} \{\text{ind}(y_j)\}. \end{array}\}$$

We claim that this language is the language obtained from w_0 by iterated hairpin completion.

First, we prove that $\mathcal{H}_\alpha^*(w_0) \supseteq R_{m1}(w_0)$. Let $w \in R_{m1}(w_0)$. By definition, any word in $R_{m1}(w_0)$ can be factorized as $w = x_s \cdots x_1 w_0 \overline{y_1} \overline{y_2} \cdots \overline{y_t}$. Compare the leftmost factor x_s and the complement of the rightmost factor $\overline{y_t}$ with respect to their index. Assume that $\text{ind}(x_s) \leq \text{ind}(y_t)$. Then $w \geq_s \overline{\alpha} \overline{y_t} \geq_s \overline{\alpha} \overline{x_s}$. Hence, one-step left hairpin completion can derive w from the word $x_{s-1} \cdots x_1 w_0 \overline{y_1} \cdots \overline{y_t}$. In the case when $\text{ind}(x_1) > \text{ind}(y_t)$, the same argument implies that $w \in \mathcal{RH}_\alpha(x_{s-1} \cdots x_1 w_0 \overline{y_1} \cdots \overline{y_{t-1}})$. Due to $\max_{1 \leq i \leq s} \{\text{ind}(x_i)\} \leq \max_{1 \leq j \leq t} \{\text{ind}(y_j)\}$, the repetition of this process eventually reduces w_0 into a word $w_0 \overline{y_1} \cdots \overline{y_j}$ for some $1 \leq j \leq t$. Because of the condition on y_1 and our discussion above, $w_0 \rightarrow_{\mathcal{RH}} w_0 \overline{y_1} \rightarrow_{\mathcal{RH}} \cdots \rightarrow_{\mathcal{RH}} w_0 \overline{y_1} \cdots \overline{y_j}$ is valid. Thus, $w \in \mathcal{H}_\alpha^*(w_0)$.

Secondly, we prove the opposite inclusion by induction on the length of derivation by hairpin completion. Clearly $w_0 \in L(w_0)$. Let us assume that a word in $\mathcal{H}_\alpha^*(w_0)$ can be written as $x_s \cdots x_1 w_0 y_1 \cdots y_t$ with $\max_{1 \leq i \leq s} \{\text{ind}(x_i)\} \leq \max_{1 \leq j \leq t} \{\text{ind}(y_j)\}$. Let $j = \max_{1 \leq j \leq t} \{\text{ind}(y_j)\}$. If left hairpin completion extends this word to the left by x , then $\overline{\alpha} \overline{y_1} \cdots \overline{y_t} \geq_s \overline{\alpha} \overline{x}$ and this means $x \in \{u_1, \dots, u_j\}^+$ (see Lemma 1). Thus, there exist $x_{s'}, \dots, x_{s+1} \in \{u_1, \dots, u_j\}$ such that $x = x_{s'} \cdots x_{s+1}$ and $\max\{\text{ind}(x_{s'}), \dots, \text{ind}(x_{s+1}), \text{ind}(x_s), \dots, \text{ind}(x_1)\} \leq j$. It is trivial that this inequality remains valid in the right hairpin completion.

Theorem 3. *For any $m \geq 1$ and a non-crossing $(m, 1)$ word $w_0 \in \alpha \Sigma^* \overline{\alpha}$, the language $\mathcal{H}_\alpha^*(w_0)$ is regular.*

The key idea in the above discussion is that if a word in $\mathcal{H}_\alpha^*(w_0)$ begins with the longest α -prefix u_m of w_0 , then hairpin completion can extend it to the right by any of α -prefix of w_0 . This idea has a broader range of applications. Let $w_0 \in \alpha \Sigma^* \cap \Sigma^* \overline{\alpha}$ be a non-crossing (m, n) -word for some $m, n \geq 1$ with $\text{Pref}_\alpha(w_0) = \{u_1, \dots, u_m\}$ and $\text{Suff}_{\overline{\alpha}}^*(w_0) = \{\overline{v_1}, \dots, \overline{v_n}\}$. Proposition 3 says that if $u_m = v_n$, then $\text{Suff}_{\overline{\alpha}}^*(w_0) = \overline{\text{Pref}_\alpha^*(w_0)}$. For $m \geq 2$, the rightmost occurrence of α on w_0 does not overlap with the suffix $\overline{\alpha}$ of w_0 (Lemma 7). Thus, $\mathcal{H}_\alpha^*(w_0) = \{u_1, \dots, u_m\}^* w_0 \{\overline{u_1}, \dots, \overline{u_m}\}^*$.

Corollary 3. *Let $w_0 \in \alpha \Sigma^* \cap \Sigma^* \overline{\alpha}$ be a non-crossing (m, n) -word. If $u_m = v_n$, then $\mathcal{H}_\alpha^*(w_0)$ is regular.*

4.2 Iterated hairpin completion of $(2, 2)$ non-crossing words

In contrast to the result obtained in the previous subsection, Example 1 shows that there exists an $(m, 2)$ non-crossing word whose iterated hairpin completion is non-regular with $m = 3$. This result motivates the study of $(2, 2)$ non-crossing words reported here. Let $w_0 \in \alpha\Sigma^* \cap \Sigma^*\bar{\alpha}$ be a non-crossing $(2, 2)$ -word. We can employ Corollary 2 to see that $\mathcal{H}_\alpha(w_0) = \{w_0, v_2 w_0, w_0 \bar{u}_2\}$. This further implies that the suffix $\bar{\alpha}$ of any word in $\mathcal{H}_\alpha^*(w_0)$ can bind with the second α on the (unique) factor w_0 on the word for right hairpin completion.

Let us define the following regular language:

$$R_{22L} = v_2^*(v_2 w_0) \bar{v}_2^* \cup (v_2^+ u_2)^* v_2^* (v_2 w_0) \bar{v}_2^* (\bar{u}_2 \bar{v}_2^+)^+.$$

We will show that this language is exactly the set of words obtained by iterated hairpin completion from $v_2 w_0$.

In order to prove that $\mathcal{H}_\alpha^*(v_2 w_0) \supseteq R_{22L}$, it suffices to present the following process:

$$\begin{aligned} v_2 w_0 &\xrightarrow{^*_{\mathcal{RH}}} v_2 w_0 \bar{v}_2^{j_0} \\ &\xrightarrow{\mathcal{RH}} v_2 w_0 \bar{v}_2^{j_0} \bar{u}_2 \bar{v}_2 \\ &\xrightarrow{^*_{\mathcal{RH}}} v_2 w_0 \bar{v}_2^{j_0} \bar{u}_2 \bar{v}_2^{j_1} \\ &\xrightarrow{\mathcal{RH}} v_2 w_0 \bar{v}_2^{j_0} \bar{u}_2 \bar{v}_2^{j_1} \bar{u}_2 \bar{v}_2 \\ &\xrightarrow{^*_{\mathcal{RH}}} v_2 w_0 \bar{v}_2^{j_0} \bar{u}_2 \bar{v}_2^{j_1} \dots \bar{u}_2 \bar{v}_2^{j_{t-1}} \bar{u}_2 \bar{v}_2 \\ &\xrightarrow{^*_{\mathcal{LH}}} v_2^{i_0} v_2 w_0 \bar{v}_2^{j_0} \bar{u}_2 \bar{v}_2^{j_1} \dots \bar{u}_2 \bar{v}_2^{j_{t-1}} \bar{u}_2 \bar{v}_2 \\ &\xrightarrow{\mathcal{LH}} v_2 u_2 v_2^{i_0} v_2 w_0 \bar{v}_2^{j_0} \bar{u}_2 \bar{v}_2^{j_1} \dots \bar{u}_2 \bar{v}_2^{j_{t-1}} \bar{u}_2 \bar{v}_2 \\ &\xrightarrow{^*_{\mathcal{LH}}} v_2^{i_1} u_2 v_2^{i_0} v_2 w_0 \bar{v}_2^{j_0} \bar{u}_2 \bar{v}_2^{j_1} \dots \bar{u}_2 \bar{v}_2^{j_{t-1}} \bar{u}_2 \bar{v}_2 \\ &\xrightarrow{^*_{\mathcal{LH}}} v_2^{i_s} u_2 \dots v_2^{i_1} u_2 v_2^{i_0} v_2 w_0 \bar{v}_2^{j_0} \bar{u}_2 \bar{v}_2^{j_1} \dots \bar{u}_2 \bar{v}_2^{j_{t-1}} \bar{u}_2 \bar{v}_2 \\ &\xrightarrow{^*_{\mathcal{RH}}} v_2^{i_s} u_2 \dots v_2^{i_1} u_2 v_2^{i_0} v_2 w_0 \bar{v}_2^{j_0} \bar{u}_2 \bar{v}_2^{j_1} \dots \bar{u}_2 \bar{v}_2^{j_{t-1}} \bar{u}_2 \bar{v}_2^{j_t}. \end{aligned}$$

Next, we prove the opposite inclusion by induction on the length of derivation by hairpin completion from $v_2 w_0$. Obviously, $v_2 w_0 \subseteq R_{22L}$. Assume that all words obtained from $v_2 w_0$ by at most n -times hairpin completion are in R_{22L} . Let w_n be such a word and consider a word w_{n+1} such that $w_n \xrightarrow{\mathcal{H}} w_{n+1}$. Consider the case when this hairpin completion is right one. The rightmost occurrence of α on w_n is the second α on its (unique) factor w_0 . Therefore, if we let $w_{n+1} = w_n \bar{\alpha}$ and then $x\alpha \leq_p (v_2^+ u_2)^* v_2^* v_2 u_2$. Since u_2 and \bar{v}_2 are the respective shortest nonempty α -prefix and $\bar{\alpha}$ -suffix of w_0 , Lemma 2 implies that $x \in (v_2^+ u_2)^* v_2^*$. Note that R_{22L} is closed under catenating a word in $(v_2^+ u_2)^* v_2^*$ to the right. Thus, $w_{n+1} \in R_{22L}$. The case when $w_n \xrightarrow{\mathcal{LH}} w_{n+1}$ can be proved in a symmetric manner.

Due to the symmetry of u_2 and \bar{v}_2 , we can easily construct a regular language R_{22R} which is equivalent to $H_\alpha^*(w_0 \bar{u}_2)$. Now the regularity of $\mathcal{H}_\alpha^*(w_0)$ has been proved.

Theorem 4. *For a $(2, 2)$ non-crossing word $w_0 \in \alpha\Sigma^*\bar{\alpha}$, the language $\mathcal{H}_\alpha^*(w_0)$ is regular.*

4.3 Iterated hairpin completion of $(3, 2)$ non-crossing words

Theorem 4 and Example 1 motivate our investigation of non-crossing $(3, 2)$ words. Actually, Theorem 5, a main contribution of this paper, provides a characterization of the regularity of iterated hairpin completion of a non-crossing $(3, 2)$ -word in terms of the commutativity of the α -prefixes and $\bar{\alpha}$ -suffixes of the word.

Let $w_0 \in \alpha\Sigma^* \cap \Sigma^*\bar{\alpha}$ be a non-crossing $(3, 2)$ -word (so $\alpha \neq \bar{\alpha}$) with $\text{Pref}_\alpha(w_0) = \{\lambda, u_2, u_3\}$ and $\text{Suff}_{\bar{\alpha}}(w_0) = \{\lambda, \bar{v}_2\}$. Note that u_2 (v_2) must be primitive; otherwise, its primitive root is also an α -prefix (resp. $\bar{\alpha}$ -suffix) of w_0 and w_0 would not be a $(3, 2)$ -word any more. As a result, u_2 commute with v_2 (u_3) if and only if $u_2 = v_2$ (resp. $u_3 = u_2^2$). Recall also that $u_3 \neq v_2$ must hold for w_0 to be $(3, 2)$ -word (Proposition 3). Thus, if u_3 and v_2 commute, then $u_3 = v_2^2$ and $u_2 = v_2$. In other words, the commutativity between u_3 and v_2 is reduced to the commutativity between u_2 and u_3 and the commutativity between u_2 and v_2 , and hence, not essential.

Corollary 2 states that $\mathcal{H}_\alpha(w_0) = \{w_0\} \cup \{v_2w_0, w_0\bar{u}_2, w_0\bar{u}_3\}$. Let us ask the question of whether iterated hairpin completion can generate a same word from $w_0\bar{u}_2$ and $w_0\bar{u}_3$. We partially answer this question in a broader setting for arbitrary $m \geq 3$ and $n \geq 1$.

Lemma 9. *Let $w_0 \in \alpha\Sigma^* \cap \Sigma^*\bar{\alpha}$ be a non-crossing (m, n) -word for some $m \geq 3$ and $n \geq 1$ with $\text{Pref}_\alpha(w_0) = \{u_1, \dots, u_m\}$. For integers i, j with $1 < i < j$, if $u_j \in \{u_2, \dots, u_{j-1}\}u_i$, then $\mathcal{H}_\alpha^*(w_0\bar{u}_j) \subseteq \mathcal{H}_\alpha^*(w_0\bar{u}_i)$; otherwise, $\mathcal{H}_\alpha^*(w_0\bar{u}_j) \cap \Sigma^*w_0\bar{u}_i\Sigma^* = \emptyset$.*

Proof. Let $u_j = xu_i$ for some $x \in \{u_2, \dots, u_{j-1}\}$. Lemma 7 implies that $w_0\bar{u}_i \rightarrow_{\mathcal{RH}} w_0\bar{u}_i \bar{x} = w_0\bar{u}_j$ is possible. Thus, the inclusion holds. Conversely, if the intersection is not empty, then Theorem 1 implies that $\bar{\alpha} \bar{u}_j = \bar{\alpha} \bar{u}_i \bar{y}$ for some $y \in \Sigma^+$. Then, due to Lemma 3, this equation gives $y \in \{u_2, \dots, u_{j-1}\}$; thus, $u_j \in \{u_2, \dots, u_{j-1}\}u_i$. \square

We can employ Lemma 9 in our current setting of non-crossing $(3, 2)$ -words to observe that if $u_3 = u_2^2$, then $\mathcal{H}_\alpha^*(w_0\bar{u}_3) \subseteq \mathcal{H}_\alpha^*(w_0\bar{u}_2)$; otherwise, $\mathcal{H}_\alpha^*(w_0\bar{u}_3) \cap \Sigma^*w_0\bar{u}_2\Sigma^* = \emptyset$. Thus, for example, if $u_3 \neq u_2^2$, then $\mathcal{H}_\alpha^*(w_0\bar{u}_3) \cap \mathcal{H}_\alpha^*(w_0\bar{u}_2) = \emptyset$.

In this subsection, we first prove that the commutativity of u_2 with v_2 or with u_3 is a sufficient condition for $\mathcal{H}_\alpha^*(w_0)$ to be regular.

Lemma 10. *If $u_2 = v_2$, then the language $\mathcal{H}_\alpha^*(w_0)$ is regular.*

Proof. Let $w_0 = w\bar{v}_2$ for some $w \in \alpha\Sigma^* \cap \Sigma^*\bar{\alpha}$. Observe that w is a non-crossing $(3, 1)$ -word with u_2, u_3 being its nonempty α -prefix. Lemma 7 implies that $|u_2| + 2|\alpha| < |w|$, which means that hairpin completion can extend w to the right by \bar{v}_2 and result in w_0 . If $|u_3| + 2|\alpha| \leq |w|$, then hairpin completion can also generate $w\bar{u}_3$, but it is not essential in the following discussion whether this is possible or not. Let us consider only the case when it is possible. Then $\mathcal{H}_\alpha^*(w)$, which is regular due to Theorem 3, is $\{w\} \cup \mathcal{H}_\alpha^*(w\bar{u}_2) \cup \mathcal{H}_\alpha^*(w\bar{u}_3)$. As we have seen above, if $w\bar{u}_3 \in \mathcal{H}_\alpha^*(w)$, then either $\Sigma^*w\bar{u}_2\Sigma^* \cap \mathcal{H}_\alpha^*(w\bar{u}_3) = \emptyset$ or

$\mathcal{H}_\alpha^*(w\bar{u}_2) \supseteq \mathcal{H}_\alpha^*(w\bar{u}_3)$. In any case, $\mathcal{H}_\alpha^*(w_0) = \mathcal{H}_\alpha^*(w) \cap \Sigma^* w\bar{u}_2 \Sigma^*$, and hence, is regular. \square

Now it is easy to see that $\mathcal{H}_\alpha^*(w_0)$ is regular when u_3 commutes with v_2 . Since w_0 is $(3, 2)$ -word, v_2 must be primitive and u_3 is equal to either v_2 or v_2^2 . In the former case, u_2 is a proper prefix of v_2 so that w_0 has \bar{u}_2 and would not be a $(3, 2)$ -word. Thus, the latter must be the case. In this case, the prefix v_2 of u_3 , which is the primitive root of u_3 , is an α -prefix of w_0 (Lemma 5), and hence, in order for w_0 to be a $(3, 2)$ -word, $u_2 = v_2$ must hold, and this brings the conclusion according to Lemma 10.

Lemma 11. *If $u_3 = u_2^2$, then the language $\mathcal{H}_\alpha^*(w_0)$ is regular.*

Proof. Lemma 10 makes it sufficient to consider the case when u_2 does not commute with v_2 . Since $\mathcal{H}_\alpha^*(w_0) = \{w_0\} \cup \mathcal{H}_\alpha^*(v_2 w_0) \cup \mathcal{H}_\alpha^*(w_0 \bar{u}_2) \cup \mathcal{H}_\alpha^*(w_0 \bar{u}_2^2)$ (when the reader check this, recall Lemma 7) and $\mathcal{H}_\alpha^*(w_0 \bar{u}_2) \supseteq \mathcal{H}_\alpha^*(w_0 \bar{u}_2^2)$, we will show the regularity of the second and third terms of this equation and that is enough for our purpose.

First, we prove that $\mathcal{H}_\alpha^*(w_0 \bar{u}_2)$ is regular. Let $w_0 = u_2 w$, where $w \in \alpha \Sigma^* \cap \Sigma^* \bar{\alpha}$ is a $(2, 2)$ -word with $\text{Pref}_\alpha(w) = \{\lambda, u_2\}$ and $\text{Suff}_{\bar{\alpha}}(w) = \{\lambda, \bar{v}_2\}$. We can easily check that

$$\mathcal{H}_\alpha^*(w) = \{w, w\bar{u}_2, w\bar{u}_2^2\} \cup \mathcal{H}_\alpha^*(u_2 w \bar{u}_2) \cup \mathcal{H}_\alpha^*(u_2 v_2 w \bar{u}_2) \cup \mathcal{H}_\alpha^*(v_2 w).$$

As done in the proof of Lemma 10, the non-commutativity between u_2 and v_2 implies that $(\mathcal{H}_\alpha^*(u_2 v_2 w \bar{u}_2) \cup \mathcal{H}_\alpha^*(v_2 w)) \cap \Sigma^* u_2 w \Sigma^* = \emptyset$. Thus, $\mathcal{H}_\alpha^*(w) \cap \Sigma^* u_2 w \Sigma^* = \mathcal{H}_\alpha^*(w_0 \bar{u}_2)$. Since w is a non-crossing $(2, 2)$ -word so that $\mathcal{H}_\alpha^*(w)$ is regular (Theorem 4), and hence, so is $\mathcal{H}_\alpha^*(w_0 \bar{u}_2)$.

Next, we prove the regularity of $\mathcal{H}_\alpha^*(v_2 w_0)$. We can let $w_0 = w' \bar{v}_2$ for some $(3, 1)$ -word w' . This means that $v_2 w'$ is a $(4, 1)$ -word with $\text{Pref}_\alpha(v_2 w') = \{\lambda, v_2, v_2 u_2, v_2 u_2^2\}$ and the empty $\bar{\alpha}$ -suffix. Thus,

$$\mathcal{H}_\alpha^*(v_2 w') = \{v_2 w'\} \cup \mathcal{H}_\alpha^*(v_2 w' \bar{v}_2) \cup \mathcal{H}_\alpha^*(v_2 w' \bar{u}_2 \bar{v}_2) \cup \mathcal{H}_\alpha^*(v_2 w' \bar{u}_2^2 \bar{v}_2).$$

Using the essentially same argument as above, we obtain $\mathcal{H}_\alpha^*(v_2 w') \cap \Sigma^* v_2 w' \bar{v}_2 \Sigma^* = \mathcal{H}_\alpha^*(v_2 w_0)$. Since the iterated hairpin completion of non-crossing $(4, 1)$ -word is regular (Theorem 3), $\mathcal{H}_\alpha^*(v_2 w')$ is regular and so is $\mathcal{H}_\alpha^*(v_2 w_0)$.

Combining what have been proved in the previous two paragraphs together, we conclude the regularity of $\mathcal{H}_\alpha^*(w_0)$. \square

To summarize the results obtained so far, any of two of the α -prefixes and the complements of $\bar{\alpha}$ -suffixes of w_0 , i.e., u_2, u_3, v_2 , must not commute in order for $\mathcal{H}_\alpha^*(w_0)$ not to be regular.

Lemma 12. *If $u_3 = u_2 v_2$, then the language $\mathcal{H}_\alpha^*(w_0)$ is regular.*

Proof. Due to Lemma 10, it suffices to consider this problem under the assumption $u_2 \neq v_2$, which is equivalent to that u_2 does not commute with v_2 under our problem setting.

We have $\mathcal{H}_\alpha^*(w_0) = \{w_0\} \cup \mathcal{H}_\alpha^*(v_2 w_0) \cup \mathcal{H}_\alpha^*(w_0 \overline{u_2}) \cup \mathcal{H}_\alpha^*(w_0 \overline{v_2} \overline{u_2})$. As done before, we will check that the second, third, and fourth terms of the union above are regular. The regularity of the third one is from $\text{Pref}_\alpha(w_0 \overline{u_2}) = \{\lambda, u_2, u_2 v_2\}$ and $\text{Suff}_{\overline{\alpha}}(w_0 \overline{u_2}) = \{\lambda, \overline{u_2}, \overline{v_2} \overline{u_2}\}$ and Corollary 3.

In order to check that the second term is regular, let $w_0 = w_1 \overline{v_2}$, where w_1 is a $(3, 1)$ -word. Then $v_2 w_1$ is a $(4, 1)$ -word, and

$$\mathcal{H}_\alpha^*(v_2 w_1) = \{v_2 w_1\} \cup \mathcal{H}_\alpha^*(v_2 w_1 \overline{v_2}) \cup \mathcal{H}_\alpha^*(v_2 w_1 \overline{u_2} \overline{v_2}) \cup \mathcal{H}_\alpha^*(v_2 w_1 \overline{v_2} \overline{u_2} \overline{v_2}).$$

Since $v_2 w_1 \overline{v_2} \rightarrow_{\mathcal{RH}} v_2 w_1 \overline{v_2} \overline{u_2} \overline{v_2}$ and $\mathcal{H}_\alpha^*(v_2 w_1 \overline{u_2} \overline{v_2}) \cap \Sigma^* v_2 w_1 \overline{v_2} \Sigma^* = \emptyset$, we have $\mathcal{H}_\alpha^*(v_2 w_0) = \mathcal{H}_\alpha(v_2 w_1 \overline{v_2}) = \mathcal{H}_\alpha^*(v_2 w_1) \cap \Sigma^* v_2 w_1 \overline{v_2} \Sigma^*$. The regularity of $\mathcal{H}_\alpha^*(v_2 w_1)$ is due to Theorem 3 so that $\mathcal{H}_\alpha(v_2 w_0)$ is regular.

What remains to be considered is the fourth term. One can let $w_0 \overline{v_2} \overline{u_2} = u_2 v_2 w_2$ for some non-crossing $(1, 4)$ -word w_2 . Then $\mathcal{H}_\alpha^*(w_2) = \{w_2\} \cup \mathcal{H}_\alpha^*(u_2 w_2) \cup \mathcal{H}_\alpha^*(u_2 v_2 w_2) \cup \mathcal{H}_\alpha^*(u_2 v_2^2 w_2)$ holds, and we can easily see that $\mathcal{H}_\alpha^*(u_2 v_2 w_2) = \mathcal{H}_\alpha^*(w_2) \cap \Sigma^* u_2 v_2 w_2 \Sigma^*$. The regularity of $\mathcal{H}_\alpha^*(w_0 \overline{v_2} \overline{u_2})$ was proved. \square

Theorem 5. *Let $w_0 \in \alpha \Sigma^* \cap \Sigma^* \overline{\alpha}$ be a non-crossing $(3, 2)$ -word with $\text{Pref}_\alpha(w_0) = \{\lambda, u_2, u_3\}$ and $\text{Suff}_{\overline{\alpha}}(w_0) = \{\lambda, \overline{v_2}\}$. Then $\mathcal{H}_\alpha^*(w_0)$ is regular if and only if one of the following three conditions holds:*

1. u_2 commutes with v_2 ;
2. u_2 commutes with u_3 ;
3. $u_3 = u_2 v_2$.

Proof. Let $R = u_3 u_2^{\geq 2} v_2 w_0 \overline{u_2}^{\geq 2} \overline{u_3}$, which is a regular language. Under the assumption that none of the conditions 1-3 holds, $L := \mathcal{H}_\alpha^*(w_0) \cap R = \{u_3 u_2^i v_2 w_0 \overline{u_2}^i \overline{u_3} \mid i \geq 2\}$ holds. As mentioned previously, if the second condition does not hold, which is equivalent to $u_3 \neq u_2^2$, then $HC_\alpha^*(w_0 \overline{u_3})$ cannot contain any word in the above intersection. Thus, $L = (\mathcal{H}_\alpha^*(w_0 \overline{u_2}) \cap R) \cup (\mathcal{H}_\alpha^*(v_2 w_0) \cap R)$. Using Lemmas 3 and 4, we can easily prove the emptiness of the second intersection of the above sum. This check is left to the reader, and the authors recommend them to check at least $\mathcal{H}_\alpha^*(v_2 w_0 \overline{u_3} \overline{v_2}) \cap R = \emptyset$ because this check involves the important fact that $\overline{\alpha} \overline{u_2} \leq_p \overline{\alpha} \overline{u_3}$ implies $u_3 = u_2^2$ and causes a contradiction. As a result, we have $L = \mathcal{H}_\alpha^*(w_0 \overline{u_2}) \cap R$. Informally speaking, in order to produce a word in R from w_0 , we first have to extend w_0 to the right by $\overline{u_2}$.

Now we can extend $w_0 \overline{u_2}$ to the right by $\overline{u_2}$ i -times to obtain $w_0 \overline{u_2}^i$. If this obtained word is extended to the left, then the word will be in $u_2 \Sigma^* w_0 \Sigma^* \overline{u_2}$. Let us check that $u_2 \Sigma^* w_0 \Sigma^* \overline{u_2} \cap u_3 \Sigma^* w_0 \Sigma^* \overline{u_3} = \emptyset$. If the intersection is not empty, then $u_3 \alpha \leq_p u_2 x \alpha$ for some $x \in \{u_2, u_3, v_2\}^+$. Due to Lemma 2, $u_3 \in u_2 \{u_2, v_2\}^+$, but actually we can say $u_3 \in u_2 \{u_2, v_2\}$ for u_3 is the second shortest nonempty α -prefix of w_0 . However, this means that either the condition 1 or 2 holds, and contradicts our assumption. Thus, we have only one choice; extending $w_0 \overline{u_2}^i$ to the right by $\overline{u_3}$.

As mentioned above, $\overline{\alpha} \overline{u_2} \leq_p \overline{\alpha} \overline{u_3}$ cannot hold so that we cannot extend $w_0 \overline{u_2}^i \overline{u_3}$ further to the right to obtain a word in R . Thus, we should extend this

word to the left either by $u_3u_2^j$ for some $j \leq i$ or by $u_3u_2^i v_2$. Lemmas 3 and 4 prove that the former choice will not lead us to any word in R . Now it suffices to mention that extending $u_3u_2^i v_2 w_0 \overline{u_2}^i \overline{u_3}$ further to the left because such an extension force the contradictory relation $\overline{\alpha} \overline{u_2} \leq_p \overline{\alpha} \overline{u_3}$ to hold. \square

5 Conclusion

In this paper, we focused on finding conditions that a word $w_0 \in \alpha\Sigma^* \cap \Sigma^*\overline{\alpha}$ must satisfy so that its iterated hairpin completion $\mathcal{H}_\alpha^*(w_0)$ is a regular language. We classified the set of all non-crossing words according to the number m of occurrences of α and the number n of occurrences of $\overline{\alpha}$ on a given word. For the cases when $n = 1$ and when $m = n = 2$, we proved that the iterated hairpin completion of a non-crossing (m, n) -word is regular. We also found a necessary and sufficient condition under which the iterated hairpin completion of a non-crossing $(3, 2)$ -word is regular. This approach can be generalized to arbitrary non-crossing (m, n) -words, with the cases $(m, 1)$ and $(2, 2)$ being the induction base of an inductive proof. Future works include considering the same problem for crossing-words. In this case, Lemma 7 or Theorem 1 does not hold any more, and hence, it may get harder to analyze the derivation processes of how a word is obtained from a given word w_0 by iterated hairpin completion. In addition, we investigated only the case when the suffix of length k of an initial word w_0 is the complement of its prefix of the same length, but we eventually have to consider w_0 in $\alpha\Sigma^* \cap \Sigma^*\overline{\beta}$, where β might not be equal to α (double-primer hairpin completion). We can easily observe that one-step hairpin completion with respect to α (β) derives a word in $\beta\Sigma^* \cap \Sigma^*\overline{\beta}$ (resp. $\alpha\Sigma^* \cap \Sigma^*\overline{\alpha}$) from w_0 . Thus, results obtained in this study of single-primer hairpin completion are important step towards this most general setting of the regularity test problem of iterated hairpin completion of a single word. Another direction of research is to consider stopper sequences as in Whiplash PCR [7, 20].

References

- [1] Leonard M. Adleman. Molecular computation of solutions to combinatorial problems. *Science*, 266(5187):1021–1024, 1994.
- [2] M. Arita and S. Kobayashi. DNA sequence design using templates. *New Generation Computing*, 20:263–277, 2002.
- [3] D. Cheptea, C. Martín-Vide, and V. Mitrana. A new operation on words suggested by DNA biochemistry: Hairpin completion. In J.-G. Dumas, editor, *Transgressive Computing*, pages 216–228, 2006.
- [4] C. Choffrut and J. Karhumäki. Combinatorics of words. In G. Rozenberg and A. Salomaa, editors, *Handbook of Formal Languages*, volume 1, pages 329–438. Springer-Verlag, Berlin-Heidelberg-New York, 1997.

- [5] V. Diekert and S. Kopecki. Complexity results and the growths of hairpin completions of regular languages (extended abstract). In M. Domaratzki and K. Salomaa, editors, *Proc. of CIAA 2010*, volume 6482 of *LNCS*, pages 105–114. Springer, 2011.
- [6] N. J. Fine and H. S. Wilf. Uniqueness theorem for periodic functions. *Proceedings of the American Mathematical Society*, 16(1):109–114, 1965.
- [7] M. Hagiya, M. Arita, D. Kiga, K. Sakamoto, and S. Yokoyama. Towards parallel evaluation and learning of boolean μ -formulas with molecules. In H. Rubin and D. Wood, editors, *Proc. of 3rd DIMACS Workshop on DNA Based Computers*, volume 48 of *DIMACS Series in Discrete Mathematics*, pages 57–72, 1999.
- [8] M. Ito, P. Leupold, F. Manea, and V. Mitrana. Bounded hairpin completion. *Information and Computation*, 209(3):471–485, 2011.
- [9] N. Jonoska, D. Kephart, and K. Mahalingam. Generating DNA codewords. *Congressus Numerantium*, 156:99–110, 2002.
- [10] N. Jonoska and K. Mahalingam. Languages of DNA based code words. In J. Chen and J. Reif, editors, *Proc. of DNA Computing 9*, volume 2943 of *LNCS*, pages 61–73. Springer, 2004.
- [11] A. Kameda, M. Yamamoto, A. Ohuchi, S. Yaegashi, and M. Hagiya. Unravel four hairpins! *Natural Computing*, 7:287–298, 2008.
- [12] L. Kari, S. Konstantinidis, E. Losseva, P. Sosík, and G. Thierrin. A formal language analysis of DNA hairpin structures. *Fundamenta Informaticae*, 71(4):453–475, 2006.
- [13] L. Kari and K. Mahalingam. Watson-Crick conjugate and commutative words. In *Proc. of DNA 13*, volume 4848 of *LNCS*, pages 273–283, 2008.
- [14] Steffen Kopecki. On the iterated hairpin completion. Technical report, Computer Science, University of Stuttgart, 2010.
- [15] F. Manea, C. Martín-Vide, and V. Mitrana. On some algorithmic problems regarding the hairpin completion. *Discrete Applied Mathematics*, 157:2143–2152, 2009.
- [16] F. Manea, V. Mitrana, and T. Yokomori. Two complementary operations inspired by the DNA hairpin formations: Completion and reduction. *Theoretical Computer Science*, 410(4-5):417–425, 2009.
- [17] F. Manea, V. Mitrana, and T. Yokomori. Some remarks on the hairpin completion. *International Journal of Foundations of Computer Science*, 21(5):859–872, 2010.

- [18] R. Okazaki, T. Okazaki, K. Sakabe, K. Sugimoto, and A. Sugino. Mechanism of DNA chain growth, i. possible discontinuity and unusual secondary structure of newly synthesized chains. *Proceedings of the National Academy of Sciences of the United States of America*, 59(2):598–605, 1968.
- [19] G. Păun, G. Rozenberg, and T. Yokomori. Hairpin languages. *International Journal of Foundations of Computer Science*, 12(6):837–847, 2001.
- [20] K. Sakamoto, D. Kiga, K. Komiya, H. Gouzu, S. Yokoyama, S. Ikeda, H. Sugiyama, and M. Hagiya. State transitions by molecules. *Biosystems*, 52(1-3):81–91, 1999.
- [21] M. Takinoue and A. Suyama. Molecular reactions for a molecular memory based on hairpin DNA. *Chem-Bio Informatics Journal*, 4:93–100, 2004.
- [22] M. Takinoue and A. Suyama. Hairpin-DNA memory using molecular addressing. *Small*, 2(11):1244–1247, 2006.
- [23] K. S. Wilson and P. H. von Hippel. Transcription termination at intrinsic terminators: The role of the RNA hairpin. *Proceedings of the National Academy of Sciences of the United States of America*, 92(19):8793–8797, 1995.